SQ2

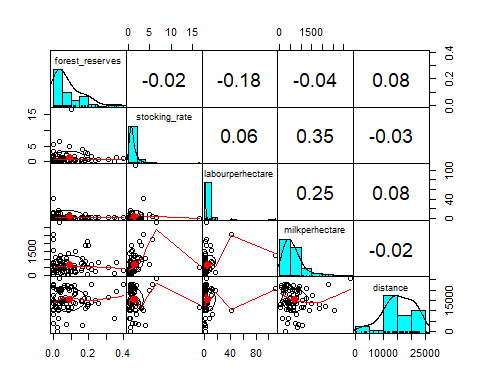
## Results of SQ2: Do farmers with more intensive systems have relatively more forest reserves?

### Loading R packages and dataframe

library(tidyverse)  
library(ggplot2)  
library(psych)  
library(heplots)  
load("Dataframes/si.df.rda")

### Scatterplots, histograms and correlations of forest reserves, stocking rates, annual milk yields per hectare, annual labour days per hectare and distances to highway

model.var <- c(12,32, 36, 37, 14) # forest\_reserves,stocking rate, milkperhectare, labourperhectare, distance  
pairs.panels(si.df[,model.var],  
 gap=0,  
 bg=c("red")[si.df$SPS],  
 pch = 21)

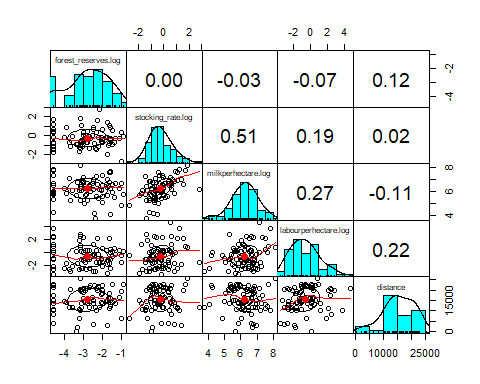


### Log-transformation of forest reserves

si.df$forest\_reserves.log <- log(si.df$forest\_reserves+0.01) ## +0.01 to avoid infinitifs

### Scatterplots, histograms and correlations of log-forest reserves, log-stocking rates, log-annual milk yields per hectare, log-annual labour days per hectare and distances to highway

log.var <- c(44,41:43,14) # forest\_reserves.log,stocking\_rate.log , milkperhectare.log, labourperhectare.log, distance  
pairs.panels(si.df[,log.var],  
 gap=0,  
 bg=c("red", "green")[si.df$SPS],  
 pch = 21)

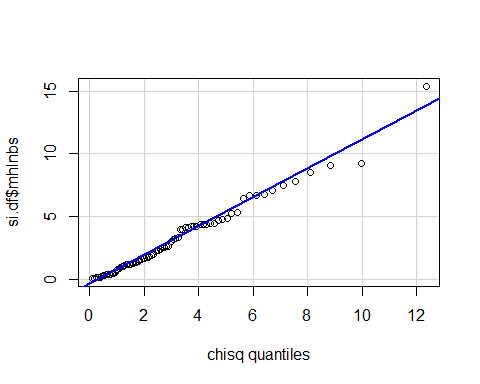


### Calculating Mahalanobis distance of log-stocking rates, log-milk yields and log-labour days

si.df<-si.df[!is.na(si.df$forest\_reserves.log),] # remove NA in forest\_reserves.log  
si.mhlnbs\_outl <- mahalanobis(si.df[,c(41:43)],colMeans(si.df[,c(41:43)]),cov(si.df[,c(41:43)]))  
# Calculated Mahalanobis distances  
si.df$mhlnbs <- round(si.mhlnbs\_outl,3) #Add to dataframe

###Q-Q plot of Mahalanobis distance and theoretic Chi-square quantiles

qqPlot(si.df$mhlnbs, distribution = "chisq", df=3, envelope = FALSE, id=FALSE)



### Outlier definition as logic variable in dataframe

outlier <- 5.5  
si.df$outlier\_mhlnbs <- ifelse(si.df$mhlnbs > outlier, TRUE, FALSE)  
summary(si.df$outlier\_mhlnbs)

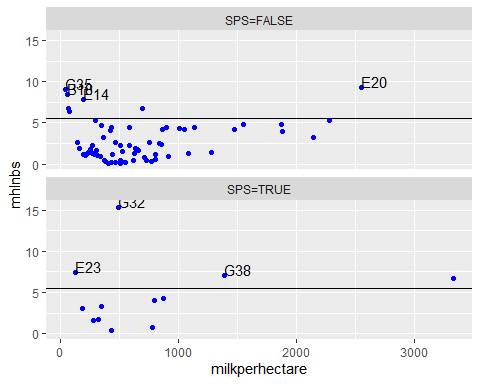
## Mode FALSE TRUE   
## logical 69 11

### Key variables of identified outliers

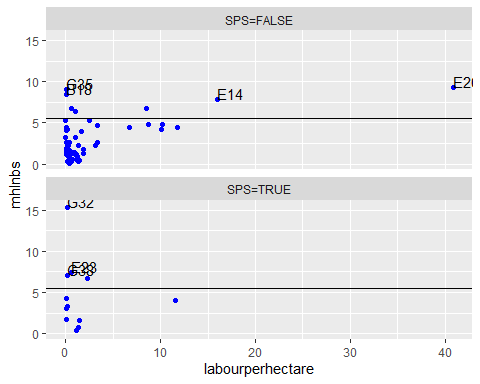
## id\_house SPS stocking\_rate labourperhectare milkperhectare mhlnbs  
## 1 B18 FALSE 0.59707602 0.07000000 57.63158 8.491  
## 2 B25 TRUE 6.45410628 2.34782609 3332.60870 6.744  
## 3 B28 FALSE 0.31827957 1.06451613 70.64516 6.407  
## 4 B30 FALSE 0.44444444 0.64242424 66.36364 6.704  
## 5 E14 FALSE 0.25964444 15.99879997 197.10000 7.823  
## 6 E15 FALSE 4.92753623 8.49826060 698.26087 6.700  
## 7 E20 FALSE 3.44444444 40.84500000 2555.00000 9.236  
## 8 E23 TRUE 0.06349206 0.63340659 128.35165 7.455  
## 9 G32 TRUE 16.72380952 0.19047619 486.66667 15.386  
## 10 G35 FALSE 0.24567281 0.04020101 44.02010 9.048  
## 11 G38 TRUE 0.23386243 0.19047619 1390.47619 7.062

### Scatterplots of Mahalanobis distance and milk yields, labour days and stocking rates

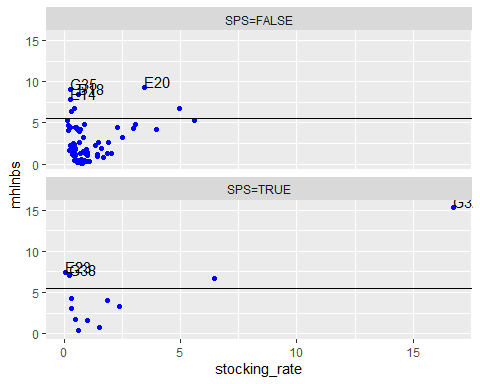
labs <- c("FALSE"="SPS=FALSE", "TRUE"="SPS=TRUE")  
par(mfrow=c(3,1))  
ggplot(data = si.df, mapping = aes(x=mhlnbs, y= milkperhectare))+  
 geom\_point(color="blue")+  
 geom\_vline(xintercept = outlier)+  
 facet\_wrap(~SPS, dir = "v", labeller = labeller(SPS=labs))+  
 coord\_flip()+  
 geom\_text(aes(x=mhlnbs, y= milkperhectare, label=ifelse(mhlnbs>7,as.character(id\_house),'')),hjust=0,vjust=0)



ggplot(data = si.df,mapping = aes(x=mhlnbs, y= labourperhectare))+  
 geom\_point( color="blue")+  
 geom\_vline(xintercept = outlier)+  
 facet\_wrap(~SPS, dir = "v", labeller = labeller(SPS=labs))+  
 coord\_flip()+  
 geom\_text(aes(x=mhlnbs, y= labourperhectare, label=ifelse(mhlnbs>7,as.character(id\_house),'')),hjust=0,vjust=0)



ggplot(data = si.df,mapping = aes(x=mhlnbs, y= stocking\_rate))+  
 geom\_point( color="blue")+  
 geom\_vline(xintercept = outlier)+  
 facet\_wrap(~SPS, dir = "v", labeller = labeller(SPS=labs))+  
 coord\_flip()+  
 geom\_text(aes(x=mhlnbs, y= stocking\_rate, label=ifelse(mhlnbs>7,as.character(id\_house),'')),hjust=0,vjust=0)



### Removing outliers from dataframe

si.df <- filter(si.df, outlier\_mhlnbs==FALSE)

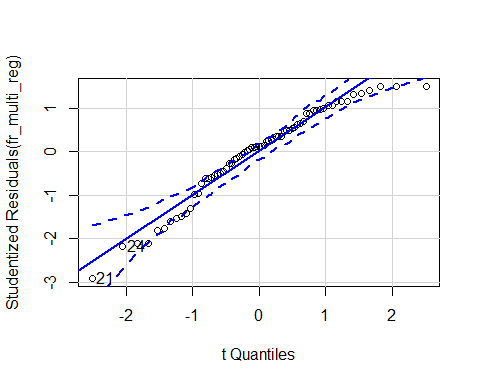
### Multiple linear regression model, dependent variable: log-forest reserves, independent variables: log-stocking-rates, log-milk yields, log-labour days, SPS, distance to highway

fr\_multi\_reg <- lm(forest\_reserves.log ~ stocking\_rate.log+milkperhectare.log+labourperhectare.log+SPS+distance, data = si.df)  
summary(fr\_multi\_reg)

##   
## Call:  
## lm(formula = forest\_reserves.log ~ stocking\_rate.log + milkperhectare.log +   
## labourperhectare.log + SPS + distance, data = si.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.6089 -0.5270 0.1216 0.7052 1.5487   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.143e+00 1.697e+00 -2.441 0.0175 \*  
## stocking\_rate.log 1.117e-01 1.945e-01 0.574 0.5677   
## milkperhectare.log 1.121e-01 2.434e-01 0.460 0.6468   
## labourperhectare.log -2.968e-02 8.934e-02 -0.332 0.7408   
## SPSTRUE -7.252e-01 4.026e-01 -1.801 0.0764 .  
## distance 4.881e-05 2.433e-05 2.006 0.0492 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.053 on 63 degrees of freedom  
## Multiple R-squared: 0.1038, Adjusted R-squared: 0.03264   
## F-statistic: 1.459 on 5 and 63 DF, p-value: 0.216

### Q-Q plot of errorterms and theoretical quantiles

qqPlot(fr\_multi\_reg)



## [1] 21 24

### Variable selection by comparing AIC

step(fr\_multi\_reg)

## Start: AIC=12.83  
## forest\_reserves.log ~ stocking\_rate.log + milkperhectare.log +   
## labourperhectare.log + SPS + distance  
##   
## Df Sum of Sq RSS AIC  
## - labourperhectare.log 1 0.1224 69.956 10.950  
## - milkperhectare.log 1 0.2350 70.069 11.060  
## - stocking\_rate.log 1 0.3658 70.200 11.189  
## <none> 69.834 12.829  
## - SPS 1 3.5967 73.430 14.294  
## - distance 1 4.4602 74.294 15.101  
##   
## Step: AIC=10.95  
## forest\_reserves.log ~ stocking\_rate.log + milkperhectare.log +   
## SPS + distance  
##   
## Df Sum of Sq RSS AIC  
## - milkperhectare.log 1 0.1849 70.141 9.1316  
## - stocking\_rate.log 1 0.3482 70.304 9.2921  
## <none> 69.956 10.9495  
## - SPS 1 3.5618 73.518 12.3761  
## - distance 1 4.3623 74.318 13.1234  
##   
## Step: AIC=9.13  
## forest\_reserves.log ~ stocking\_rate.log + SPS + distance  
##   
## Df Sum of Sq RSS AIC  
## - stocking\_rate.log 1 0.8739 71.015 7.9860  
## <none> 70.141 9.1316  
## - SPS 1 3.8717 74.013 10.8390  
## - distance 1 4.1821 74.323 11.1278  
##   
## Step: AIC=7.99  
## forest\_reserves.log ~ SPS + distance  
##   
## Df Sum of Sq RSS AIC  
## <none> 71.015 7.9860  
## - SPS 1 3.7371 74.752 9.5247  
## - distance 1 3.8951 74.910 9.6704

##   
## Call:  
## lm(formula = forest\_reserves.log ~ SPS + distance, data = si.df)  
##   
## Coefficients:  
## (Intercept) SPSTRUE distance   
## -3.366e+00 -7.310e-01 4.348e-05

### AIC-optimized model, independent variables: SPS, distance to highway

fr\_sps\_distance\_reg <- lm(forest\_reserves.log ~ SPS + distance, data = si.df)  
summary(fr\_sps\_distance\_reg)

##   
## Call:  
## lm(formula = forest\_reserves.log ~ SPS + distance, data = si.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.2400 -0.4756 0.1313 0.7646 1.5256   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.366e+00 3.644e-01 -9.237 1.65e-13 \*\*\*  
## SPSTRUE -7.310e-01 3.922e-01 -1.864 0.0668 .   
## distance 4.348e-05 2.285e-05 1.903 0.0615 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.037 on 66 degrees of freedom  
## Multiple R-squared: 0.08861, Adjusted R-squared: 0.06099   
## F-statistic: 3.209 on 2 and 66 DF, p-value: 0.0468

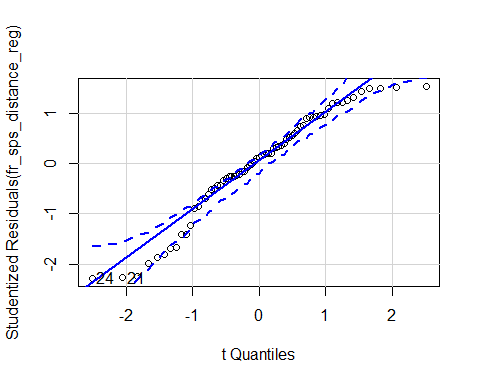
### Leaving out single observations to see reflection on estimators

head(dfbeta(fr\_sps\_distance\_reg), n=10L)

## (Intercept) SPSTRUE distance  
## 1 0.0174352876 -0.0156254842 -1.086302e-07  
## 2 0.0142237193 -0.0109759961 -1.949388e-07  
## 3 0.0034911363 -0.0035297981 2.320605e-09  
## 4 0.0387364945 -0.0226684591 -9.644552e-07  
## 5 0.0220584842 -0.0223410769 1.696212e-08  
## 6 -0.0310743995 0.0314724957 -2.389501e-08  
## 7 0.0079509393 -0.0173771386 5.657908e-07  
## 8 -0.0037058362 0.0027592341 5.681811e-08  
## 9 -0.0009173393 0.0005357745 2.290275e-08  
## 10 0.0184703077 -0.0165615861 -1.145676e-07

### Q-Q plot of errorterms of AIC optimized linear model and theoretical quantiles

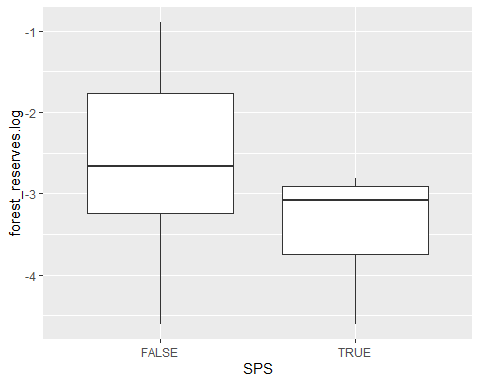
qqPlot(fr\_sps\_distance\_reg)



## [1] 21 24

### Boxplots of log-forest reserves of farmers without and with SPS

ggplot(data = si.df, mapping=aes(y=forest\_reserves.log, x=SPS))+  
 geom\_boxplot()



### One-sided t-test of mean log-forest reserves of farmers without and with SPS

fr\_ttest <- t.test(x=si.df$forest\_reserves.log[si.df$SPS==FALSE],  
 y=si.df$forest\_reserves.log[si.df$SPS==TRUE], alternative = "greater"  
 )  
fr\_ttest

##   
## Welch Two Sample t-test  
##   
## data: si.df$forest\_reserves.log[si.df$SPS == FALSE] and si.df$forest\_reserves.log[si.df$SPS == TRUE]  
## t = 2.4577, df = 13.252, p-value = 0.01424  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.1829738 Inf  
## sample estimates:  
## mean of x mean of y   
## -2.720532 -3.372854

### Transformation of means back to level values

exp(fr\_ttest$estimate)-0.01

## mean of x mean of y   
## 0.05583974 0.02429162